政策血缘关系网络构建与传播演化机理研究 *

刘 刚，李 涛，路彩霞，崔一冬

(哈尔滨工程大学 计算机科学与技术学院, 哈尔滨 150001)

摘 要：政策体系是一个庞大的系统，R 政策碎片化日益严重导致政策不一致性。本文通过挖掘政策血缘关系对政策系统进行建模，将数量庞大的政策形式化为政策血缘网络。在此基础上构建 PCL 模型，通过机器学习确定模型中节点的容量定义，实现政策网络中脆性点挖掘。该方法通过挖掘政策间的相关概念，将政策体系形式化为政策谱系树模型，再由政策谱系树的节点对之间发生"树影相吸"，最终形成基于政策血缘关系的政策血缘网络。通过政策血缘网络的传播演化机理，构建了 PCRL 模型，通过机器学习对容量参数进行参数训练，挖掘传播演化机理。最后选取政策实例进行仿真实验，验证了方法的有效性。

关键词：政策体系；政策血缘网络；机器学习；脆性点

中图分类号：P315.69　　doi: 10.3969/j.issn.1001-3695.2017.09.0918

Text lineage network construction and evolution mechanism of communication

Liu Gang, Li Tao, Lu Caixia, Cui Yidong

(*College of Computer Science & Technology*, *Harbin Engineering University*, *Harbin* 150001, *China*)

Abstract: In this paper, the lineage from a policy text perspective modeling approach. The method is an industry policy text formalized policies lineage tree. Between policy text lineage tree node "shadows attract", eventually forming policy based network text lineage relationship. And on the basis of the policy text on the basis of lineage network nodes in the network referral value, evaluate and dominant policy points found kinship network text, through the dissemination of policy text lineage network evolution mechanism constructed PCRL model, and the model based on PCL above, to achieve a brittle point of excavation and to mine the evolution mechanism. Finally, the policy text experiment in the field of modeling and analysis methods used in the field of policy text points specific area of policy research.

Key Words: policy system; policy text crucial points; Machine learning; brittle points mining

0 引言

随着我国社会主义现代化建设的不断深入发展和社会主义法治社会的不断完善，各地政策层出不穷，中国特色社会主义法律体系基本形成[1]。但是著名民法学家孙宪忠指出我国社会主义法律体系的建成仅仅表明我国已基本解决了"有法可依"的问题，如何将这些海量的政策法规体系化、完善化，迫切地需要研究人员们进一步地深入研究政策之间的关系，形成解读政策的研究工具，使我国的政策体系科学化，立法体系化。单纯通过人工的纯感性或半理性的政策分析方式，恐怕很难给出精确的结论。研究政策之间关系的存在与否、强度如何和传播机理，具有极为重要的意义。

1）国内外研究现状

现有的公共政策仿真大多采用仿真模型作为一种定量分析的分析方法。文献[13]在对政策研究的相关文献的研究基础上，可以看出，截止 2012 年 5 月 27 日，我国现有的政策仿真方法研究中，采用系统动力学作为政策仿真方法的文献有 366 篇，排在第二位的是多主体仿真方法 55 篇，数值仿真方法相关文献 30 篇。从政策的血缘关系角度对政策进行研究，现在国内外尚无相关研究成果发表。但是作为生物学领域的一个重要分支的血缘关系在数据血缘领域的研究在国外已有许多研究成果发表，值得本文借鉴。

Karvounarakis 等人研究出一种基于元组的、半环溯源的语言，这种语言支持数据查询的存储、处理和索引方案，并且通过实验证明了该语言在数据溯源上可用。Bao 等人[7]提出一种紧凑的、高效的可达性标签方案用于解答给定规范下的工作流的数据溯源。即使工作流的结构更复杂，以及由于并行和循环执行导致的任意大的规范，运用源于固定规范的事实文中展示

基金项目：黑龙江省博士后科研启动基金资助项目（LBH-Q15031）；黑龙江省教育科学规划课题（GLC121507）

作者简介：刘刚（1976-），男，黑龙江哈尔滨人，副教授，博士，主要研究方向为深度学习、大数据处理（389083838@qq.com）；李涛（1992-），男，硕士研究生，主要研究方向为自然语言处理；路彩霞（1989-），女，硕士，主要研究方向为政策网络、自然语言处理；崔一冬（1993-），女，硕士研究生，主要研究方向为自然语言处理、数据血缘.

了一次运行的紧凑的可达性标记的有效计算。标签方案使用对数长度的标签，以线性时间运行，并在常数时间内，回答任意的可达性查询。

2）本文主要研究内容

在前期研究中已经提出了一种新的政策形式化方法，即从政策血缘的角度将政策之间的血缘关系划分为政策显性血缘关系以及政策隐性血缘关系，并提出了挖掘政策隐性血缘关系的算法。

本文将在前期研究工作的基础上，进行更进一步的深化。将以政策形式化的方法搭建政策血缘网络，并在此基础上进行政策脆性点提取，进而实现政策关键节点的发现。

a)构建政策血缘网络模型

本文在前期研究工作的基础上，以政策中政策之间的关系为基础，以政策篇章为节点，政策篇章之间的相似度为政策血缘网络中边的权值，构建政策血缘网络，进而实现政策血缘网络的可视化。

b)基于政策血缘网络的 PCL 模型

研究适合于政策血缘网络基因传播方式和政策血缘网络上相继故障的负载-容量模型 PCL〔Policy Capacity Load〕模型，观察政策基因在政策血缘网络上的传播模式，探究政策的相继故障方法，构建政策血缘网络的 PCL 模型。

c)基于政策血缘网络的网络脆性点发现

随机移除政策血缘网络中的节点或者节点集，记录被移除节点或节点集的原始负载在网络中重新分配的过程，记录相继故障结束后，网络中剩余节点的数量，计算节点的脆性度，并依据节点脆性度发掘政策血缘网络中的脆性点。

1　相关工作与理论基础

在政策血缘关系的研究过程中，运用到政策网络的相关理论知识。在政策脆性点发现的研究中采用了相继故障的思想。

1.1　政策血缘网络上的相继故障现象

相继故障模型大致可分为以下三类，即负载—容量模型，耦合映像格子模型及其他相继故障模型。其中负载-容量模型又可以分为只考虑节点负荷的节点动态模型、只考虑网络中边负荷的边动态模型以及既考虑节点负荷又考虑边负荷的节点与边的混合动态模型[10]。本研究中，网络中的节点之间因血缘关系而相连，节点之间血缘关系强的，边的权值就大，反之，则小。由于用汉语描述的政策政策，可以用形似或相近的词汇替换后，使得政策政策还保持原来的语义，那么政策血缘网络的研究就适合采用负载-容量模型。

2　政策血缘网络模型构建

政策血缘网络模型的构建是政策显性要点分析和政策脆性点发现的基础。将行业政策形式化为一棵政策谱系树。政策谱系树的节点对之间发生“树影相吸”，最终形成基于政策血缘关系的政策血缘网络。

2.1　政策血缘关系挖掘

在本文中，政策之间的血缘关系用政策政策之间的相似度来度量。本文选择向量空间模型 VSM〔Vector Space Model〕来计算政策政策之间的相似度。假设政策政策 D_1 与政策政策 D_2 可分别表示为向量 $D_1=(W_1,W_2,\ldots W_i,\ldots W_k,\ldots,W_n)$ 与政策政策向量 $D_2=(W_i',\ldots W_k',X_1,\ldots X_j,\ldots X_m)$，其中 W_i 为特征项 t_i 的权重，W_i' 为政策政策 D_2 中与政策政策 D_1 相同的特征项 t_i 在政策政策 D_2 中的权重，X_j 为政策政策 D_2 中不同于政策政策 D_1 的特征项 t_j 的对应权重，n 和 m 分别为政策政策 D_1 中与政策政策 D_2 中的特征项数目，那么政策政策 D_1 与政策政策 D_2 的血缘关系可以用式（1）来衡量。

$$I_{\text{相似度}}(D1,D2)=\frac{\sum w_i \times w_i'}{|D_1|\times|D_2|} \tag{1}$$

2.2　政策谱系树的构建

2.2.1 政策谱系树的构建原理

以政策谱系树的形式形象地表示了政策之间的血缘关系。以国宪法作为根节点，紧随宪法之后的各行业基本法为第一层子节点，之后各个行业的具体政策构成其基本法的子孙节点，由此形成树状的政策体系表示被称为政策谱系树。政策谱系树[11]传递的是政策之间的显性血缘关系，称通过政策谱系树显性地传播的政策的血缘关系为政策显性血缘。

基于政策谱系树的概念，下面首先给出政策谱系树的形式化描述如下：

定义 1 政策谱系树〔policy pedigree tree〕。Ps 是一个政策集合，基于 Ps 的政策谱系树是一个二元组 PPTtree=<T,R>，有：

a）PPTtree 是一棵政策谱系树；

b）T 是特征相同的数据元素的集合，　$T \in Ps$；

c）如果|T|=0，那么 $R=\varnothing$ ，此时的 PPTtree 是一棵空的政策谱系树；

d）如果|T|=1，那么 $R=\varnothing$，PPTtree 只有一个根节点；

e）如果|T|>2，那么 R 是 T 上的一个二元关系的集合。除根节点外，对于集合 T 中的每一个节点，在关系 R 中有且仅有一个前趋节点，但可以有多个后继节点

2.2.2 政策谱系树的构建与更新算法

基于定义 2 给出的政策谱系树的模型，可以继续给出政策谱系树的构建和更新算法。

输入：政策谱系树根节点 *root*，新引入的政策节点 *newnode*，阈值 0.1。

输出：更新后的政策谱系树的根节点 *newroot* 生成新引入的政策节点 newnode 的向量模型。

if *root*= Φ　then

{　root←newnode;

newroot←root;}

else {

计算节点 *root* 与新引入的政策节点 *newnode* 的政策相似度值　$I_{\text{相似度}}(root,newnode)$;

比较阈值与 $I_{相似度}(root,newnode)$;

if $I_{相似度}(root,newnode)$>=阈值 then

将新引入节点 *newnode* 插入 *root* 节点的子节点序列中；

end if;

end if;

}

返回更新后的政策谱系树的根节点 *newroot*。

2.3　树影相吸现象

现在整个法律法规体系由原来的一棵政策谱系树，分化为包含多个行业法规体系树林立的政策森林[11]。但都衍生自宪法，各棵树之间仍隐性地保持着政策森林的统一性。在政策森林中，政策谱系树之间形成隐性血缘关系，表现为一棵树的部分节点投影到另一棵树的部分节点上，从某一方向观察两棵树部分重合，这种现象称为"树影相吸"现象。

定义 2　对于政策谱系树 A 和 B，树 A 中含有的政策节点为 $\{a1,a2,...ai,...,an\}$，树 B 中含有的政策节点为 $\{b1,b2,...bj,...,bm\}$，若存在满足 $I_{相似度}(a_i,b_j)\in(0,1]$，其中 $0<i<n,0<j<m$ 关系的一对或者多对节点对，则称树 A 与树 B 发生了"树影相吸"。如图 1 所示

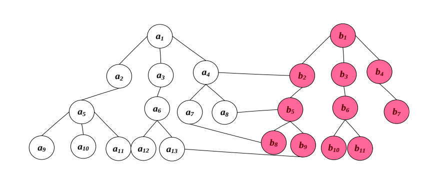


图 1　政策谱系树 A 与政策谱系树 B 发生"树影相吸"之后的示意图

2.4　树影相吸的构造过程

单节点对的连接极有可能增加节点的负载，进而影响血缘关系在森林中传递的畅通性，当节点的负载超过其容量时，将影响政策谱系树的连通性，引起与节点相连的政策父节点和政策子节点的过载，使节点故障进一步扩大，最终导致整个政策体系的崩溃。

为了避免单节点对的连接导致的政策体系的崩溃，本文有选择地选择节点对，使得既可以适应政策血缘关系的描述，又不增加政策森林的复杂度。政策谱系树子树对间连接算法具体过程如下：

输入：子树 A，子树 B。

输出：更新后的政策模型 A'。

获取子树 A 中的相吸节点集 a_i;

获取子树 B 中的相吸节点集 b_j;

创建变量 temp[i];

for 子树 A 中的每一个节点 a_i{

for 子树 B 中的每一个节点 b_j

{利用式(1)计算节点 a_i 与节点 b_j 的相似度值 $I_{相似度}(a_i,b_j)$;

将 $I_{相似度}(a_i,b_j)$存入变量 *temp[i]*；}

}

对 *temp[i]*中的值进行降序排列；

选取前 *k* 个节点对；

for i=1 to k {

将节点 b_j 插入节点 a_i 的子节点序列中；

}

返回更新后的政策模型 A'。

其中 1<=n<=N,1<=m<=M，其中 N，M 分别是子树 A 与子树 B 中的节点总数。

2.5　政策血缘网络的构建

在政策森林中，政策隐性基因的传播，导致了"树影相吸"现象的发生，并且这种现象经常发生于不同的政策谱系树对之间，使得政策森林中的政策谱系树变得不再是相互独立的树木，而是充斥着不同强度的血缘关系的网络。本文称这种由于政策基因的传播而形成的政策关系网络为政策血缘网络，并将政策抽象为网络中的节点，政策之间的关系抽象为连接连个政策节点的网络中的边。

定义 3　利用图 G=(V, E, W)代表政策血缘网络，其中，V 是政策政策节点的集合，即 V={V_j|V_j 代表政策政策节点}；E 是无向边的集合，即 E={（V_i ,V_j）|单节点对政策政策 V_i 与政策政策 V_j 之间存在血缘关系}，W 为无向边的权值，值的大小代表政策政策之间的血缘关系的强弱。

3　基于政策血缘网络的脆性点发现

本章在政策血缘网络的基础上,研究政策血缘网络基因传播方式,通过构建政策血缘网络上相继故障的负载—容量模型 PCL（policy capacity load）模型，发掘政策血缘网络中的脆性点。

3.1　政策脆性点发现思想

在政策血缘网络中，如果某项关键政策的修改或删除导致与之相关的政策的正常执行受阻，那么影响会随着政策之间的血缘关系逐渐向相邻的政策节点传播,从而造成局部政策瘫痪。

定义 4　脆性点。设政策血缘网络 G 由 n 个节点 V_i（$i=1,2,\ldots,n$）组成，如果存在一个节点或者节点集，当节点因受到外部攻击而发生故障，从而使与之相连的节点和边也发生故障，有

$$\lim_{t\to\infty}\delta(G)=\infty \tag{2}$$

则称政策血缘网络 G 具有脆性，该节点或者节点集就是政策血缘网络的脆性点或脆性点集。

3.2　构建 PCL 模型

本文构建的基于政策血缘网络的 PCL 模型是节点容量和初始负载不相关的模型。

定义 5　称句子 S 中是否存在词语 x，句子 S 语义的变化程度为词语 x 在句子 S 中的贡献度[11]。

词语对句子的贡献度用词语的权值来计算。假设句子 S 中含有词语为 $\{t_1,t_2,\cdots,t_n\}$，词语 $t_i(i=1,2,\cdots,n)$ 的权值为 v_i，

则词语 t_i 对句子 S 的贡献值 $ConWd_{is}$ 为

$$ConWd_{is} = \frac{v_i}{\sum_{i=1}^{n} v_i} = v_i * (\sum_{i=1}^{n} v_i)^{-1} \qquad (3)$$

定义 6　称句子 S 对含有句子 S 的段落 Chp 的语义的影响程度称为句子 S 在段落 Chp 中的贡献度[11]。如果段落 P 中含有的子句序列为 $\{s_1, s_2, \cdots, s_n\}$，子句 $s_i (i = 1,2,\cdots,n)$ 的权值设为 sv_i，那么子句 s_i 在段落 P 中的贡献值 $ConSen_i$ 的计算公式如下：

$$ConSen_i = \frac{sv_i}{\sum_{i=1}^{n} sv_i} = sv_i * (\sum_{i=1}^{n} sv_i)^{-1} \qquad (4)$$

政策政策中的概念信息存储在政策概念信息表 Information of Concept 中，Information of Concept 的表结构为{规范词；规范词的权值；词对句子的贡献度[11]；词所在句子的编号；句子所在篇章的编号；句子对段落的贡献度[11]}。采用公式 W_i=计算政策篇章中概念在政策篇章中的权值.

定义 5　政策血缘网络中节点 i 在时刻 t 的负载为

$$I_i(t) = \sum_{i=1}^{n} \alpha W_i(t), 1 \le i \le n \qquad (5)$$

其中：n 为 t 时刻时政策政策节点 i 中含有的概念数量，α 为可变参数。$I_i(0) = \sum_{i=1}^{m} \alpha W_i(0), 1 \le i \le m$（初始负载）

定义 6　定义网络中节点 i 的容量 $C_i = \sqrt{n}$ 其中 n 为政策篇章向量中含有的概念的数量。

定义 7　在政策血缘网络的负载均衡过程中，以网络边权为比例均衡到相邻节点 j 上，那么邻居节点 j 在 t 时刻的额外负载为

$$\Delta I_j(t) = \frac{sim(i,j) \times I_i(t)}{\sum_{h=1}^{k} sim(i,j)} = \frac{sim(i,j)}{\sum_{h=1}^{k} sim(i,j)} \times I_i(t), j \in \theta \qquad (6)$$

其中：θ 是政策政策节点 i 的邻居节点集，k 为节点 i 的相邻节点数目。Sim(i,j)是政策政策节点 i 和政策政策节点 j 之间的相似度值。当节点 i 只有一个相邻节点时，节点 i 发生故障时，原来附加在节点 i 上的负载将全部分配给节点 i 的邻居节点。

3.3　脆性点挖掘

3.3.1 相继故障过程

本文为研究政策网络中碎片化政策的影响，基于政策血缘网络构建容量与负载正相关的容量-负载模型研究政策网络的相继故障过程，具体过程如下：对网络进行随机攻击和蓄意攻击，删除网络中的节点 i，该节点的负载按照相关度成正比分配给与之相邻的节点集，计算与之相邻的节点 j 的负载，如果负载大于节点容量，则节点 j 发生故障，设为故障节点进行负载均衡过程，如果网络中所有节点的负载都小于节点的容量，那么相继故障结束。演化过程如图 2 所示。

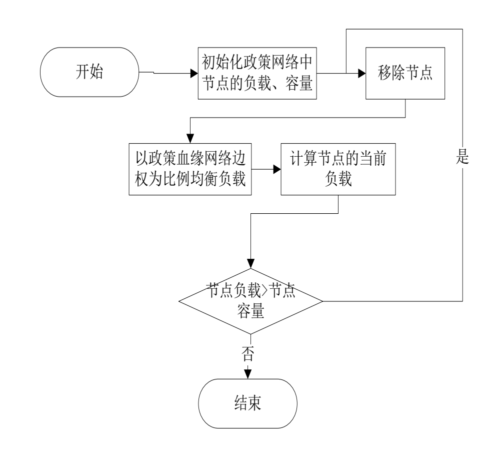


图 2　PCL 模型演化过程

3.3.2 脆性点挖掘

在政策血缘网络中，如果移除的节点是政策血缘网络中的“重要节点 i”，那么节点 i 的移除将触发政策血缘网络的相继故障现象，并将降低网络的连通性。网络中移除的节点越重要，网络遭到破坏的程度越严重，也即相继故障结束后，网络中节点的数量越少，因此可以用网络中节点的数目来衡量网络中节点的重要度。

定义 8　节点脆性度。在含有 N 个节点的政策血缘网络 G=（V,E,W）中，由于节点 i 的移除触发的网络中发生故障的节点的数量为 n，则称 I_{ki} 为节点 i 的脆性度。

$$I_{ki} = \frac{n}{N} \qquad (7)$$

显然 $0 \le I_{ki} \le 1$。I_{ki} 值越大，节点受到外界攻击时越容易触发网络中的其他节点发生故障。

在定义 10 的基础上给出政策血缘网络中节点的脆性度评估的具体算法如下：

输入：政策血缘网络 G 的邻接矩阵 G，概念信息表 *InformationofConcept*

输出：节点脆性度 I_{ki}

a）定义存储政策血缘网络中每个节点的初始负载的数组 *L[ ]*；存储每个节点容量的数组 *C[ ]*；

b）定义数组变量 *temp[]*，　*temp*，*sum*；//统计故障节点的个数

//计算政策血缘网络中节点的初始负载和容量

c）*for i=1 to n* {计算节点 i 的初始负载 $I_i(0)$；//*n* 为政策血缘网络中节点　的数目

d）$L[i] \leftarrow I_i(0)$；

e）$C[i] \leftarrow (1+a) * I_i(0)$；　}

/*对变量 *temp* 赋值，变量 *temp* 中始终存放网络中因过载而删除的节点*/

f）temp $\leftarrow i$；

g）删除节点 *temp*；/*提取节点 *temp* 的相邻节点集 θ */

h）遍历节点 *temp* 的相邻节点 *j*；

i）将节点 j 存储于数组 *temp[]* 中；

j）*while*（$|\theta|$）{//负载均衡

k) for θ 中的每一个节点 j{计算节点 temp 的相邻节点 j 在 t 时刻的总负载 $I_j(t)$；

l）if $I_j(t) \le C_j$ then 节点 j 不发生故障；

m）本次负载均衡结束；

n）if $I_j(t) > C_j$ then 节点 j 变为故障节点；

o）从网络中移除节点 j 以及与节点 j 相连的边；

p） 转 （17）；

q） sum=sum+1； } }

r）用式（6）计算节点的脆性度；

3.4 PCRL 模型的构建

本文后期研究中对 PCL 的模型上进一步改进，提出容量与负载成正相关的容量—负载模型 PCRL(policy capacity positive related to load）模型。PCRL 模型改进了节点容量的定义，负载均衡策略，并且构建了基于节点相关度的政策血缘网络，在此相关度网络上按照节点间的相关度成比例分配故障节点的负载。

3.4.1 节点容量的定义

节点的初始负载为节点正常状态下政策政策规定的条款，新旧政策的冲突导致政策中部分条款遭到破坏，因此政策节点的最大承受限度与初始负成正相关，其定义如下：

定义 9　政策血缘网络中节点 i 的容量：网络中节点 i 的容量与节点 i 的初始负载 I_i 成正比，网络中节点 j 的容量 C_j 定义为

$$C_i = (1+\lambda)I_i\ ,\quad \lambda > 0 \tag{7}$$

λ 为容量参数，基于决策树 C4.5 算法计算得出，容量参数是在 0~1 之间的实数，以 0.1 为单位，对节点集进行映射，将得到的节点重要性与已知节点重要度比较，计算信息增益率，本文中取节点的信息增益率为节点负载集映射得到的数据集与已知节点重要度的测试集的逼近程度，容量参数作为待确定容量定义的属性，把信息增益率最大的属性作为决策树的根节点，分裂出子节点，判断产生的子节点是否为叶子节点，是叶子节点则完成计算，得到节点容量的定义，否则以 0.01 为单位，循环进行上述过程，确定节点容量参数，确定节点容量定义。

3.4.2 按政策相关度分配负载

政策相似度：直观表现为两政策文字的相同程度。政策相关度：直观表现为两政策内容的主体及涉及的元素的相互作用力度。相似的两政策不一定相关，相关的两政策可能相似。假设发生故障节点为 i，节点 i 上的负载 $I_i(t)$ 以节点间的相关度为比例分配到与之相连的节点 j 上，此时节点 j 的负载增量 $\Delta Ij(t)$ 为

$$\Delta Ij(t) = \frac{re(i,j)}{\sum_{h=1}^{k} re(i,j)} \times I_i(t),\quad j \in \theta \tag{8}$$

其中：$re(i,j)$ 是政策政策节点 i 和其邻节点 j 之间的相关度值。

3.4.3 构建相关性网络

1）政策政策相关性计算思想

本文的研究以政策政策中的条款为基本单位，记为政策子句。一个句子的成分由句子主体、主体的动作和动作的承受者构成，一般句子的主体和动作承受者为名词，隐含的相关性因素即为名词本身，可通过相同关键词比较明显的得到，即运用句子向量的相似性计算即可得出。句子中的动作，直观表现为动词，通常隐含的相关性因素为动作的执行者和动作的执行范围等。如图 3 所示。

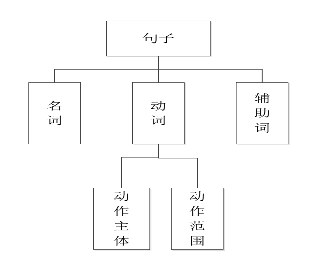


图 3　句子的构成成分

两个句子间的相关性表现为主体或承受者的相关和结构性相关。其中主体或承受者相关表现为句子中名词的相同或相近，句子结构性相关表现为由句子中动词涉及的相关因素出现交集而造成的相关。根据政策政策中的政策子句相关度计算政策政策的相关度。

2）相关度计算形式化探究

针对政策血缘网络对政策预处理过程中，引进动词相关因素集，例如监督=（监督部门，监督范围），动词相关因素集建成一个词典。词典的存储形式：关键词 → 动作的发出者 → 动作的接受者。提取关键词时，对于动词进行筛选，对照动词相关空间词典中的元素。如果关键词=元素，用动作的发出者、动作的接受者替换。得到句子的相关因素向量表示 $f(c_1, c_2 \ldots c_n)$[12]。运用向量相似度算法[13]，得到两个句子的相似度，设定阈值 *a*,相似度大于 *a*，则相关设置句子 *i* 和句子 *j* 的相关度 *r(i,j)=1* 否则 *r(i,j)=0*。相关度算法如下：

输入：子句预处理后所得词汇集合 *WS*，相关因素表 *RelatedWordTable*。

输出：子句词汇相关性因素分解后的词汇集合 *RWS*。

a)获取子句预处理后所得的词汇集合 *WS*；

b)令 $RWS = WS$；

c) *for* each *RWS* 中的词汇 *word* {

d) 获取预处理过程中对词汇 *word* 的词性标记 *sign*； //词性过滤

e) *if* *sign* 为动词{ //选择动词替换

f)在相关因素表 *RelatedWordTable* 中检索词汇 *word*；

g)*if* 词汇 *word* 在相关因素表 *RelatedWordTable* 中{

h)将 *word* 替换为相关因素表 *RelatedWordTable* 中的相关主体和相关范围;

i) }*else if* $sign$不为名词{

j) 删除词汇 *word*；}//进行词性过滤，删除辅助性词汇} }

k)返回词汇集合 *RWS*；

3)相关性网络的构建

按照上面提到的算法，计算网络中节点的相关度，依次确定网络节点的连接关系，即确定网络中的边和边的权值。基于政策子句的政策政策相关度计算的具体算法如下：

输入：政策子句的相关度矩阵 RS，节点 a 的政策子句起始序号 a_0,a_n，节点 b 的政策子句起始序号 b_0, b_m。

输出：政策政策 *a,b* 的相关度 *RE(a,b)*。

a)初始化政策政策相关度 RE(a,b)，令 RE(a,b)=0；定义变量 sum_i=0；

b)for 矩阵横坐标 i,(a_0<=i<=a_n){

c)for 矩阵纵坐标 j,(b_0<=j<=b_m)

{记录政策子句 *i* 和 *j* 的相关度 *r(i,j)*；

e) sum_i +=r(i,j); }}

f)定义变量 S(a,b)=0;S(a,b)=(a_n-a_0)*(b_m -b_0);

g)政策政策相关度 RE(a,b)= sum_i /S(a,b);

4 实验结果验证及其应用

4.1 政策血缘网络的可视化结果

本实验以社会养老保险相关政策政策为例，进行数据处理，得到政策血缘网络的节点和边的数据，用 EXCEL 将数据进行处理和转换，导入 GEPHI 软件进行运算，绘制出该政策政策的血缘关系网络，如图 4 所示

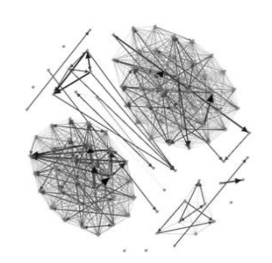


图 4 政策血缘网络

4.2 政策脆性点发现

政策脆性点发现实验由政策篇章中词汇权值计算、节点容量和负载的计算、负载均衡策略三部分组成。

4.2.1 PCRL 模型的构建的结果

改进后的容量定义：Ci=（1+a）Li，按照 4.3 中容量参数的训练算法计算 a 的值。网络中节点发生故障时，可认为政策向量中特征项均受到破坏，负载变为原来的两倍，因此节点容量参数 a 应满足 0<=a<=1,整理数据，得到效果见图 5 所示。

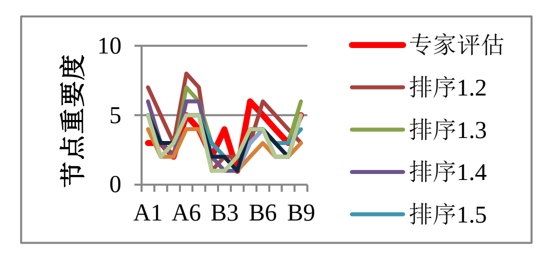


图 5 容量参数训练

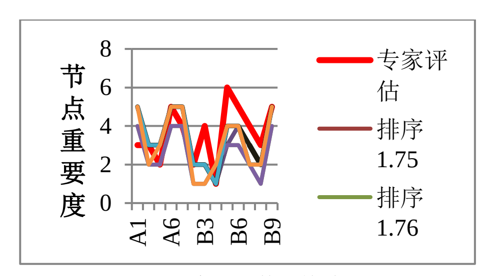


图 6 容量参数训练效果

由图 5 分析可得，猜想正确，容量参数 a=1 时，网络的相继故障不明显，节点重要性评估相差很大。B4 的初始负载是 0.8272444，并且在根节点周围以 0.01 为单位，所以以 a=0.8 为决策树根节点，循环计算，处理数据，折线图对比分析如图 6 所示。

按照算法，选取 0.81 为决策树分枝点，由图 6 可得，容量参数为 0.81 效果最好，验证了算法的正确性。取 0.01 为单位，循环计算得出结果参数为 a=0.810。容量与负载正相关的 PCRL 模型更加接近专家评估排序，对节点的重要性评估更加准确，完成了 PCL 模型的改进。改进结果如图 7 所示。

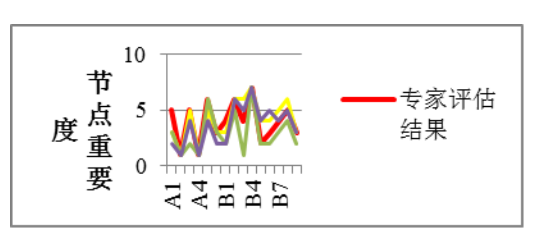


图 7 改进结果对比图

4.2.2 负载均衡过程的改进验证

负载均衡过程的改进分为两部分实现：计算政策血缘网络中有边相连的节点对的相关度，在前期研究阶段构建的基于相似性的政策血缘网络上按照节点间的相关度成比例分配负载。对政策血缘网络中有边相连的节点对的相关度进行评估，结果如表 3 所示。

4.2.3 构建基于相关性的政策血缘网络

基于节点相关性的政策血缘网络的构建分为政策政策子句间相关性计算和政策政策间相关性计算。根据算法计算得到所有政策政策间的相关度，设置节点间存在边的阈值 a 为 0.1，构建的相关性网络如图 8 所示。

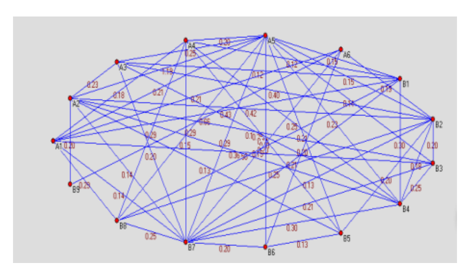


图 8 阈值为 0.1 的基于相关性的政策血缘网络

4.2.4 PCL 模型实验结果

1)政策篇章向量表示

首先对每个政策政策进行政策预处理，得到篇章中的关键词，计算每个篇章中词的权值，词对句子的贡献度，句子对篇章的贡献度。如表 1 所示，是政府采购法中第一章的部分词汇信息表。

2)节点的容量和初始负载的计算

由第 1）步计算所得每个篇章节点的概念信息为源数据，录入信息表 InformationofConcept 中，计算政策血缘网络中每个节点的初始负载和容量，计算结果如图 9 所示。

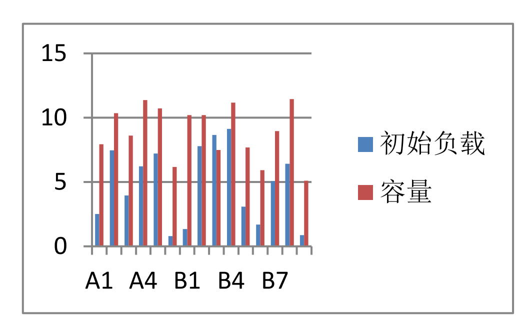


图 9　政策血缘网络中节点故障结果

3)负载均衡过程

政策血缘网络的负载均衡实验过程包括三个部分：a）随机选择要删除的政策血缘网络中的节点 i；b）节点 i 的初始负载在网络中进行负载重新分配；c）负载均衡结束后网络的情况。随机移除政策血缘网络的每个节点之后，网络中发生的相继故障过程如表 2 所示。

4)脆性点评估结果

依据上述结果，对政策血缘网络中的节点的脆性度进行评估，评估结果如图 10 所示。

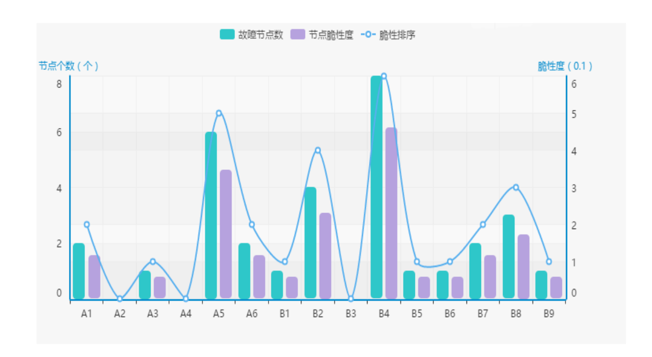


图 10　政策血缘网络脆性点评估结果

4.2.5 实验结果对比分析

本文对于脆性点的挖掘进行了两种改进，在原网络上按照节点相关性成比例分配负载和构建相关性网络并按照节点相关性成比例分配负载，本节将以专家评估结果为准，对两种改进方式的实验结果与 PCL 模型相对比，分析改进效果。

按照在相关性网络上节点间按相关度负载均衡方式即 PCRL 模型的结果中，节点的重要度评估更贴近专家评估结果，其中对于节点 B3 的重要性评估两种改进方法较 PCL 模型都明显提高。B3 是采购法第三章，规定了政府采购的方式，内

容涉及需要废标的情形、采用竞争性谈判方式采购等，与招标投标法具有很强的血缘关系。点间的相互影响出发，基于节点相关性给出节点的关系，可以更好地挖掘网络中的脆性点。

表 1　政府采购法第一章部分词汇信息表

关键词	权值	关键词对句子的贡献度	关键词所在的句子编号	关键词	权值	关键词对句子的贡献度	关键词所在的句子编号
规范	6	0.003442	1	自治区	3	0.001721	15
公共利益	3	0.001721	1	实施	6	0.002754	16
工程	3	0.001721	8	服务	3	0.001721	21
政府	1	0.000574	8	国务院	5	0.002869	21
……	……	……	……	……	……	……	……
规定	6	0.011016	9	监督	3	0.001721	22
公平竞争	3	0.001721	9	管理部门	6	0.003442	22

表 2　政策血缘网络中节点的初始负载和容量

节点	相继故障过程	网络中剩余节点
A1	A1->B3->A6	A3,A5,B1,B2,B4,B5,B6,B7,B8,B9,
A2	无	A1,A3,A5,A6,B1,B2,B3, B4,B5,B6,B7,B8,B9
A5	A5->B3->A6->B4->A1->B2->B6	A3,B1,B5,B7,B8,B9
A6	A6->B3->A1	A3,A5,B1,B2,B4,B5,B6,B7,B8,B9
B1	B1->B3	A1,A3,A5,A6,B2,B4,B5,B6,B7,B8,B9
B2	B2->B3->A1->A6->B6	A3,A5,B1,B4,B5,B7,B8,B9

表 3 基于政策血缘网络的节点相关度

节点对		相关度	节点对		相关度
A1	B1	0.430769231	A5	B1	0.148351648
A1	B2	0.07142857	A5	B3	0.136363636
A1	B3	0.5	A5	B4	0.230769231
A1	B5	0.14	A6	B1	0.038461538
A3	B1	0.11538462	A6	B9	0.333333333

5 结束语

针对海量政策法规中普遍存在的政策不统一和政策“碎片化”现象，本文从政策血缘的角度出发，提出了将政策网络、政策血缘理论以及社会网络中的相继故障理论三者相结合的研究技术路线与研究内容。根据政策体系的血缘关系，构建政策谱系树、政策森林，根据“树影相吸”现象，构建政策血缘网络模型，将海量政策建成统一网络，建立基于政策血缘网络的 PCL 模型。同时基于政策血缘网络的 PCL 模型和改进后的 PCRL 模型挖掘出政策血缘网络中的脆性点及脆性点集。

参考文献：

[1] 赵之滢，于海，朱志良，等. 基于网络社团结构的节点传播影响力分析 [J]. 计算机学报, 2015, 38 (4): 754-755.

[2] 王莉，程学旗. 在线社会网络的动态社区发现及演化 [J]. 计算机学报 2015, 38 (2): 223-224.

[3] 周涛，张子柯，陈关荣，等. 复杂网络研究的机遇与挑战 [J]. 电子科技大学学报, 2014 (1) .

[4] 胡磊. 浅谈中国特色社会主义法学理论体系研究 [J]. 法制博览, 2013 (5): 104-105

[5] 程克勤 李世伟 周健. 基于边权值的网络抗毁性评估方法 [J]. 计算机工程与应用. 2010 (35): 95-100.

[6] 胡磊. 浅谈中国特色社会主义法学理论体系研究. 法制博览. 2013 (5): 104-105.

[7] Schweitzer F, Fagiolo G, Sornette D, et al. The new challenges. science [J]. Economic Networks, 2009 (6): 422-425.

[8] 刘刚. 面向领域的软件需求一致性验证方法研究 [D]. 哈尔滨: 哈尔滨工程大学. 2008: 21-54.

[9] 刘铭，吴冲，刘远超，等. 基于特征权重量化的相似度计算方法 [J]. 计算机学报 2015, 38 (7): 1422-1423.

[10] Moreno Y, Pastor-Satorras et al. Critical load and congestion instabilities in scale-free networks [J]. Europhys. Lett. , 2003, 62: 292-298P

[11] Crucitti P, Latora V, Marchiori M. Model for cascading failures in complex networks [J]. Phys. Rev. E, 2004: 69

[12] Freeman L C. Conceptual clarification. social networks [J]. Centrality in Social Networks, 1979, 1: 215-239

[13] 刘影. 面向领域的隐形政策血缘挖掘方法研究 [D]. 哈尔滨：哈尔滨工程大学工学. 2013: 14-33.